Nonparametric Transfer Function Models with Localized Temporal Effect

John Carlo P. Daquis
University of the Philippines Diliman

A semiparametric transfer function model is proposed and estimated using the backfitting algorithm. Simulation studies indicated that the procedure provides robust estimates for the transfer function especially for short time series data. This provides a viable alternative to the parametric transfer function model that requires large number of time points to estimate a number of parameters of the model. Furthermore, in the presence of seasonality or structural change, the procedure generally yields more robust estimates of the transfer function model than the maximum likelihood estimates of the parameters of the model.

Keywords: transfer function model, semiparametric model, backfitting, mixed models

1. Introduction

Some time series data exhibit dynamic behavior that often invalidates univariate analysis. Given a single output time series $\{y_t\}$, variation can be explained by the past values (autocorrelation) and the present and past values of an input series $\{x_t\}$. For example, daily water consumption can be affected by consumption pattern in the past as well as the present weather conditions. Past weather conditions may also contribute through its inherent autocorrelation structure.

Box et al. (1994) pointed out the transfer function model that relates the output and the input series along with the autocorrelations of both the output and input series through

$$y_t = v(B)x_t + n_t. \tag{1}$$

In Equation (1), $v(B) = \sum_{j=-\infty}^{\infty} v_j B^j$ is the filter called the transfer function and η_t is the noise process independent of the input function $v(B)x_t$, and B is the backshift operator. The output series is a linear function of the current and past values of the input series. Box et al. (1994) also discussed an estimation procedure that involves two steps: identification of the transfer function and the noise model and; maximum likelihood estimation of the parameters.

There are some limitations in the transfer function model in Equation (1). While the infinite-ordered linear process can approximate both linear and non-linear functions, the model is saturated with parameters. The input and output series must also be jointly stationary and cointegrated (see Engle and Granger, 1987). Modifications of the transfer function model were proposed to address these problems. Gao (2007) used mixed modeling in order to address nonlinearity of the series. Partially linear autoregressive models were also introduced where the current value of the output y_t is a linear function of its past values and a nonlinear function on $\{x_t\}$ (see Tjøstheim et al., 2007; Li and Racine, 2006; and Gap et al., 2005, for further details). While these models addressed some of the limitations of the transfer function model, they are still vulnerable to random movements which cannot be explained by either x_t or past values of y_t, therefore the error series may still deviate from a white noise. Hidalgo (1992) proposed a linear input and nonparametric error model, while Truong and Stone (1994) considered nonparametric input and linear autoregressive error model. These semiparametric methods handle the error series well but the stationarity restrictions and other assumption in the model are often violated in short time series data.

A semiparametric form of the transfer function model is proposed which is flexible enough to capture both linear and nonlinear features of the influence of the input to the output series, i.e., the input affects the output series through a nonparametric function, while random components are added in the model to account for nonstationarity, seasonality and structural change without going through the pre-whitening methods required in the traditional parametric transfer function modeling algorithm. Then a backfitting algorithm is proposed in the estimation of the parameters and the nonparametric functions.

2. Parametric Estimation of Transfer Function Models

Suppose there are two time series $\{x_t\}$ and $\{y_t\}$ collected at the same time interval, same frequency, and are linked by some dynamic system. Box et al. (1994) as cited in Wei (2006) proposed a method of modeling such dynamics through transfer function models given in Equation (1).

The basic steps of transfer function modeling as outlined in Wei (2006) included:

1. *Pre-whiten the input series.* Transfer function models require both the input and the output series to be stationary. The pre-whitened input series is denoted by the following model:

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} x_t, \tag{2}$$

α_t is a white noise process with mean zero and variance σ^2_α. The coefficients ϕ and θ are the autoregressive and moving average parameters respectively.

2. *Transform the output series using the pre-whitened input series as the filter.* The output series will then be filtered using the pre-whitened input series defined in Step 1. The transformed output series is denoted by the following model:

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} y_t. \tag{3}$$

3. *Calculate the sample cross correlation function (CCF) between α_t and β_t and estimate the transfer function.* The CCF, via significant correlations between the input and the output series determines which lags of the input series are significantly influencing the current output value. The sample CCF therefore is vital in estimating the impulse responses and consequently, the transfer function. It is important to note that both input and output series must be pre-whitened in order to have meaningful interpretations of the CCF. If the input and output series have a rigid dependence structure and when the output series is filtered through the estimated model for the input, output is also pre-whitened.

4. *Estimate the noise series and combine it with the function in Step 4 to have the estimated transfer function model.* In transfer function modeling, the noise process is not limited to a white noise process, some general linear processes may be considered.

5. *Do diagnostic checking.* Examine the residuals and check if the residuals are indeed white noise and independent of the input series $\{x_t\}$.

The transfer function model can be viewed as a regression model on time series data. The dependent variable is the output series and the independent variables are the current and past values of the input series. The impulse response weights are the regression coefficients, which quantify the value of information from the current and past values of the input to the output.

Nonstationarity and seasonality is addressed by the pre-whitening processes and through differencing. However, shocks and interventions happen in time, and Box et al. (1994) discussed techniques in intervention analysis that can handle such time points in transfer function modeling.

3. Backfitting and Mixed Models

The backfitting procedure was proposed by Buja et al. (1989) as an iterative algorithm which uses smoothing techniques in building nonparametric regression models. Hastie and Tibshirani (1990) discussed the steps in implementing the backfitting algorithm:

1. *Initialize values*. Let $\hat{a} = \overline{Y}$ and $f_j = f_j^{(0)}, j = 1, 2, \ldots, p$.

2. *Iterative smoothing*. For repeated *j=1,2,...,p*, the smooth of x_j is obtained by smoothing the residuals on x_j. That is,

$$\hat{f}_j = smooth_j\left(y - a - \sum_{k \neq j} f_k \mid x_j\right).$$

3. *Repeat*. Step 2 is done repeatedly until the individual functions do not change. The resulting estimated model is additive:

$$\hat{y} = \hat{\alpha} + \sum_{j=1}^{p} \hat{f}_j\left(x_j\right). \tag{4}$$

The additive model has been used effectively in time series analysis. Dominici et al. (2002) used generalized additive models in their analysis of air pollution and health time series. Santos and Barrios (2012) postulated an additive model and used it in decomposing multicollinear time series. Shieke (2001) presented a nonparametric survival model based on generalized additive modeling. Asymptotic optimality of the backfitting algorithm has been established by Mammen et al. (1999) and Opsomer (2000). Hastie and Tibshirani (1990) further noted that the additive model can replace an additive decomposition of the time series model into its linear and nonlinear components.

The additive model gives a greater degree of flexibility on the postulated model since the form of the function of the input series need not be specified and the effects of each input series to the output series can be investigated individually.

While the additive model can effectively capture the dependence structure between the input and the output time series, if there are some features in y_t which are not captured by x_t (e.g. seasonality and structural change), the noise structure would consist of pure noise at some points and meaningful localized structures at some segments. This will consequently induce clustering behavior among residuals where these behaviors had been localized. Demidenko (2004) argued that when dealing with clustered data, one can use the mixed model given by

$$y_{ij} = \alpha_i + \beta x_{ij} + \varepsilon_{ij} \tag{5}$$

where α_i is the ith cluster effect, and βx_{ij} is the effect of the j^{th} observation in the i^{th} cluster.

4. Semiparametric Transfer Function Model

Alternative to the transfer function model in Equation (1), we propose a combination of nonparametric and parametric functions given by:

$$y_t = \gamma_t + \lambda_t = \sum_{i=1}^{p} g_i\left(x_{it}\right) + L_t + \varepsilon_t \quad \text{and} \quad L_t = \sum_{j=1}^{s}\left(L_{jt} + \eta_{jt}\right) \tag{6}$$

where $\{y_t\}$ is the output series, $\{x_{it}\}$ is the i[th] input series, $\{\varepsilon_t\}$ is the error series, $g_i\{x_{it}\}$ is the smooth function of y_t on x_{it}, L_t is the localized temporal effect, η_{jt} is the random variation within the j[th] cluster, p is the number of input variables, T is the length of the time series, and s is the number of clusters.

The output series is decomposed into two components, γ_t that summarizes the contributions of the inputs and the localized temporal component λ_t that accounts for all other variation in y_t (including the random shocks) that is not explained by the inputs. The contribution of the input is summarized by the additive nonparametric function $g_i(x_{it})$ that accounts for the dynamics in which the input influence the output series. No rigid assumptions on structure of the dependence of the output from the input series will be imposed, i.e., $g(\cdot)$ will be assumed to be some smooth function only. This will also avoid the tedious process of identifying the impulse response function especially for transfer functions with more than one input variables. Furthermore, the input and output series need not be stationary processes. Consequently, the absence of regularization constraints on the input and output series will cause the residuals to manifest some patterns that are not accounted by the nonparametric function $g(\cdot)$. Thus, any stylized patterns in the output not associated with the input (e.g. seasonality, trend, structural changes, etc.) will be captured by the localized temporal component λ_t. In the absence of λ_t, the noise term will manifest such perturbations as random shocks. The localized temporal effect L_t is a mixed model that will take care of other features of the time series such as seasonality, trend, or structural changes that are not necessarily associated with the input series. The noise processes (ε_t and η_{jt}) are assumed to have a zero mean and finite variance.

Equation (6) is clearly an additive model. Thus, the backfitting algorithm is used in the estimation of the nonparametric regression $g_i(\cdot)$ and the localized temporal effect L_t. The following iterative procedure is used in estimation:

Step 1: Smooth the output series based on the input series. Using some smoothing algorithms, e.g., smoothing splines, to estimate the nonparametric regressions $g_i(x_{it})$, $i = 1,2,\ldots,p$, ignoring initially the contribution of the localized temporal effects. This is the nonparametric approximation on the link between the input and the output series. The initial residual is then computed as follows:

$$y_t - \sum_{i=1}^{p} \hat{g}_i(x_t) = \delta_t \tag{7}$$

The innovation process δ_t still contains information on the temporal dependencies of the output series not explained by the input series, including the localized temporal component. The estimate $\hat{g}_i$ in Equation (7) minimizes the penalized sum of squared errors:

$$\sum_{t=1}^{T} \{y_t - g(x_t)\}^2 + \lambda \int_a^b \{g''(x)\}^2 \, dx \tag{8}$$

Step 2: Estimate the localized temporal component from the residuals δ_t in Step 1 (Equation 7). Consider the perturbations as stimulus of the clustering effect, the residuals are viewed as clustered data and the mixed model proposed by Demidenko (2004)is used. The initial error series are grouped into previously identified clusters. In the absence of *a priori* clusters, clustering can be identified from time plots, residual plots, or other sources of information like documentation of data generation, or even a quick cluster analysis of the residuals.

Step 3: Iterate from Steps 1 and 2, computing the residuals with the most recent estimates of the nonparametric function and the localized temporal effects, respectively, before the steps as implemented.

The estimation procedure takes advantage of the fact that the data generating model (6) is additive. Hence, nonparametric regression and mixed model are embedded into the backfitting algorithm to estimate the semiparametric model in Equation (6).

5. Simulation Studies

Varying scenarios accounting for different possible perturbations and characteristics of the time series data were simulated.Since transfer function models are very sensitive to the length of the time series, four different lengths of the time series are considered. The scenario for a very short time series was simulated with 36 time points, 60 time points for short time series, 180 for medium length, and 360 for a longer time series.

The output series $\{y_t\}$ is generated in two phases. First is to simulate the dependency of the output on the input series. In this case, stationarity of the input is considered, together with the form of the transfer function and the presence of a second correlated input variable. Next is to add the localized temporal effect to the output series. Localized temporal effect included (but are not limited to) seasonality, trend, and structural change towards the end of the output time series.

The data-generating model is given in Equation (9) below.

$$y_t = \frac{\sum_{i=0}^{s} \omega_i B^i}{\left(1-\sum_{j=1}^{r} \delta_j B^j\right)} B^b x_t + L_t + \varepsilon_t,\ \ L_{jt} = \sum_{j=1}^{s} L_{jt} + \eta_{jt}, \tag{9}$$

where $\varepsilon_t \sim N(0,1)$ and $\eta_{jt} \sim N(0, 0.75)$.

In simulating the input series, two scenarios were considered, a stationary AR(1) process

$$(1-0.5B)(x_t - 20) = a_t \ where\ a_t \sim N(0,1), \tag{10}$$

and a nearly nonstationary AR(1) process:

$$(1-0.99B)(x_t - 20) = a_t \ where\ a_t \sim N(0,1). \tag{11}$$

Both the stationary and nearly nonstationary series have mean 20. The goal here is to assess the performance of the proposed model under the nonstationary and nonlinear (or stochastic trend) conditions.

In Model (9), the coefficients of x_t comprise the transfer function that will capture the dependencies between the input and the output series. There are 3 parameters in the transfer function: *b*, the delay parameter that indicate the time lag until the input affects the output; *s*, the order of the numerator polynomial; and *r*, the order of the denominator polynomial. Two transfer functions are considered:

$$\left(0.25 + 0.5B^1 + 0.25B^2\right) x_{t-1} \tag{12}$$

and

$$\frac{\left(0.125 + 0.25B^1 + 0.125B^2\right)}{\left(1-0.5B^1\right)} x_{t-1} \tag{13}$$

For both transfer functions, a delay parameter b=1 is assumed, single and two input time series considered. The difference between (12) and (13) is that in (13), the output y_t is affected by one autoregressive time point.

In the case of multiple input series, Wei (2006) noted that these input series must be uncorrelated so that modeling with a single input series can be extended and applied with no difficulty, but problem arises when the two input series are correlated. One scenario will investigate the consequence of correlation existing between the two input series, i.e., simulate another time series $x_{2,t}$ which is correlated with $x_{1,t}$. This is obtained from the following:

$$x_{2,t} = 0.3x_{1,t} + e_{x1,t} \text{ where } e_{x1,t} \sim \mathrm{N}(0,3) \tag{14}$$

After simulating the dependency structure between the input and the output series, localized temporal effects are added to the output series. Seven types of localized temporal components were treated as clusters not associated with the input series but present in the output series. Note that the clusters are not necessarily contiguous time points, rather packets of time points which exhibit similar and possibly repeated patterns (structures and dependencies).

These localized temporal effects aregenerated from the following model:

$$L_t = \sum_{j=1}^{s} \left(L_{jt} + \eta_{jt} \right) \text{ where } \eta_{jt} \sim N(0, 0.75). \tag{15}$$

Assuming*s* clusters, note that this temporal effect will represent the component in the output series that is not associatedwith the input series, hence, the temporal effect will represent significant data perturbations with respect to the output series only.

Seasonality, as a temporal effect, is simulated by dividing the T time points into cycles, each of length l. The r^{th} point in each seasonal cycle is the point wherein the seasonality spikes appear, with all other time points generated as a white noise. For the trend, temporal effect is simulated by adding a constant in every cluster. The total number of time points T is divided into steps of length l. The first cluster step is generated from a white noise. The next cluster is obtained by adding a constant to the mean of the previous cluster along with noise.The nature of the structural change is also an autoregressive series of order 1.

$$(1-0.75B)(x_t-25) = a_t \text{ where } a_t \sim N(0,1) \tag{16}$$

Combinations of these localized temporal effects are also investigated. The scenarios are summarized in Table 1.

6. Results and Discussions

The predictive ability of the model is evaluated using the mean absolute prediction error (MAPE). MAPE per scenario are computed for both the parametric transfer function model in Equation (1) (called Model 1 subsequently) and the semiparametric model in Equation (6) (called Model 2 subsequently).

From Table 2, the average MAPE of Model 2 for very short time series is lower than that of Model 1. Table 3 gives the average MAPE for both Model 1 and Model 2 for the different forms of the input series across time points. Except

Table 1 Simulation Scenarios

Scenario	Description and Rationale	Notation and number of levels	Values	Descriptions
Time points	Assess robustness over varying length of time series	*t*: 4 levels	t=30 t=60 t=180 t=360	Very small Small Medium Large
Input series stationarity	Assess the effect of nonstationarity of the series	x_t: 2 levels	xt: stationary time series xt: nonstationary time series	Stationary Nonstationary
Transfer Function	Dependency of the output on the input series	$\frac{\sum_{i=0}^{s} w_i B^i}{\left(1-\sum_{j=1}^{r} w_j B^j\right)} x_{t-1}$ 2 levels	(r,s,b)=(0,2,1) (r,s,b)=(1,2,1)	Absence of the denominator terms limits the effect of the input to the output in terms of finite number of lags. Presence of denominator term in the transfer function leads to exponentially decaying contributions of infinite lags of inputs.
With a correlated 2nd input variable	Assess the effect of multiple correlated inputs	x_1+x_2: 2 levels	x_2 = 0 (single input) x_2 moderately correlated with x_1	
Localized Temporal Effect	Cluster effects not present in input series	Lt: 7 scenarios	Lt: Seasonality Lt: Step function deterministic trend Lt: Structural change towards end of the series Lt: Seasonality and Step Function Lt: Seasonality and structural change Lt: Step function and structural change Lt: Seasonality, step function and structural change	1. Annual seasonality was simulated. 2. Cluster of 12 time points, with constant increase per cluster. 3. 10% of points towards the end of input series are structurally different from the rest. Both 1 and 2. Both 1 and 3. Both 2 and 3. All 1, 2, and 3.

Table 2 Comparison of Mean Absolute Percentage Errors across Time Points

Method	36 Time Points	60 Time Points	180 Time Points	360 Time Points
Model 1	9.82992	16.65544	8.88494	10.16917
Model 2	9.19555	17.17574	10.23115	13.78936
Difference Between Model 1 and 2	0.63437	-0.52030	-1.34621	-3.62019

for the scenarios with two moderately correlated inputs series, all the average MAPE in Model 2 are lower when there are 36 time points. For short time series, Model 1 performs poorly while Model 2 exhibited robustness on the length of time series. As the length of thetime series increases, both the Models 1 and 2 yield comparable predictive abilities.

Summarized in Table 4 is the MAPE by each of the localized temporal effect across different time series lengths. Except for the case of a linear trend, Model 2 has consistently produced lower MAPE than Model 1. Again, as the length of the time series increases, the performance of Model 1improves and becomes comparable with Model 2. For output series with structural change, Model 2 yield lower MAPE across all time series lengths. There are also a number of scenarios with seasonality where Model 2 produced lower MAPE compared toModel 1. Even though the MAPEs are smaller for Model 1 for relatively longer time series, Model 2 yields comparable MAPE values.

For short time series, Model 2 yield lower MAPE in most of the scenarios. The majority of the time series with seasonality were also modeled better by Model 2. Even when there are 60 time points, it still produced better predictive ability than Model 1 for the majority of the scenarios with an exponentially decaying transfer function (presence of denominator term in the transfer function). Model 2 produced better predictive ability in fewer scenarios for time series data with 180 time points but it has consistently produced better predictions on data with seasonality or structural change.

Model 2 also addressed most of the limitations associated with Model 1. In all cases of localized temporal effects, the cross-correlation function can only identify correctly the delay lag of the input series less than 50% of the time. Except possibly for the presence of seasonality, when localized temporal effects are present in the output series, the cross-correlation function fails to specify a correct form of the transfer function. Even if a delay parameter is correctly specified, specification error is still observed.

The Model 2 in Equation (6), being nonparametric, mitigates the tedious process of identifying the model form, and therefore avoiding the problem of misspecification. Convergence in estimation is also an issue for the parametric

Table 3 Mean Absolute Percentage Errors of the Models for the Different Scenarios across Time Points

	36 Time Points		60 Time Points		180 Time Points		360 Time Points	
Method	Stationary	Nonstationary	Stationary	Nonstationary	Stationary	Nonstationary	Stationary	Nonstationary
Model 1	6.87535	14.16584	6.79417	32.62984	5.82068	14.34268	5.53185	17.80004
Model 2	6.45370	13.67491	6.87414	34.33247	6.62885	16.11554	6.19186	26.05563
Difference Between Model 1 and 2	0.42165	0.49094	-0.07997	-1.70263	-0.80817	-1.77286	-0.66001	-8.25559
Method	Finite Lags	Exponentially Decaying Lags	Finite Lags	Exponentially Decaying Lags	Finite Lags	Exponentially Decaying Lags	Finite Lags	Exponentially Decaying Lags
Model 1	11.77300	7.88684	26.12344	7.18744	12.32316	5.92116	14.87507	5.46326
Model 2	11.41788	6.97322	27.57637	6.77511	13.66234	6.79995	20.73880	6.83991
Difference Between Model 1 and 2	0.35512	0.91361	-1.45293	0.41233	-1.33919	-0.87879	-5.86373	-1.37665
Method	One Input	2 Correlated Inputs	One Input	2 Correlated Inputs	One Input	2 Correlated Inputs	One Input	2 Correlated Inputs
Model 1	9.76484	9.89500	23.17117	10.13971	8.35110	9.89322	11.55198	8.78635
Model 2	8.37146	10.01965	23.52504	10.82644	9.07457	11.38772	15.27443	12.30428
Difference Between Model 1 and 2	1.39339	-0.12465	-0.35388	-0.68673	-0.72348	-1.49450	-3.72245	-3.51793

Table 4 Mean Absolute Percentage Errors of the Models for the Different Localized Temporal Effects across Time Points

	Seasonality	Trend	Structural Change	Seasonality + Trend	Seasonality + Structural Change	Trend+ Structural Chance	Seasonality+Trend +Structural Change
36 Time Points							
Model 1	10.78124	9.37489	11.93047	11.19944	10.00243	7.32241	8.19856
Model 2	10.53586	10.34550	8.12783	10.15862	8.35691	8.37720	8.46695
Difference Between Model 1 and 2	0.24538	-0.97061	3.80263	1.04082	1.64551	-1.05478	-0.26839
60 Time Points							
Model 1	10.24794	9.54316	58.22568	10.60476	12.96061	7.01479	7.99113
Model 2	11.99423	12.40215	57.38382	11.13905	10.59880	8.45572	8.25642
Difference Between Model 1 and 2	-1.74629	-2.85899	0.84186	-0.53429	2.36182	-1.44093	-0.26529
180 Time Points							
Model 1	17.05131	7.42948	11.71852	8.09481	7.12726	6.00667	6.42706
Model 2	16.81357	10.39271	8.73840	10.55917	8.43798	8.66955	8.00664
Difference Between Model 1 and 2	0.23774	-2.96323	2.98012	-2.46437	-1.31072	-2.66288	-1.57958
360 Time Points							
Model 1	26.71893	6.20442	9.13981	11.21365	6.74314	5.09813	6.06609
Model 2	35.72079	9.98540	8.31374	17.72307	8.52897	7.82954	8.42397
Difference Between Model 1 and 2	-9.00185	-3.78099	0.82607	-6.50942	-1.78583	-2.73141	-2.35788

method (Model 1). For data sets with a finite lag transfer function, estimation in 99.80% of the datasets converged. However, for an exponentially decaying form of the transfer function, the estimation procedure (Model 1) converged only in 75.75% of the datasets. For the semiparametric transfer function procedure (Model 2), even though it invokes two-step iterative procedures, none of the datasets resulted to non-convergence.

7. Conclusion

In the presence of localized temporal effects in time series data, the parametric transfer function procedure cannot provide reliable estimates and predicted values due to the following reasons: first, it requires a rigid specification of the dependence structure between the input and output variables; second, it becomes unstable for datasets with very short time series. Furthermore, the parametric procedure is vulnerable to incorrectly specifying the form of the transfer function, particularly the order of the delay parameter. The estimation procedure for the parametric transfer function model may not converge, especially for short time series data and when the transfer function is more complicated.

On the other hand, the proposed semiparametric model is robust with respect to the length of the time series. For short series, the semiparametric model yield better predictive ability than the parametric model. As the length of the time series increases, the predictive ability of the parametric procedure improves but the semiparametric procedure still produced comparable results. Furthermore, the semiparametric model yields better predictive ability than the parametric model whenever the outputs series exhibit seasonality and/or structural change.

REFERENCES

BOX, G., M. JENKINS and G. REINSELL, 1994, *Time Series Analysis: Forecasting and Control* 3rd ed., San Francisco: Holden-Day.

BUJA, A., T. HASTIE and R. TIBSHIRANI, 1989, Linear smoothers and additive models, *The Annals of Statistics* 17(2): 453-510.

DOMINICI, F., A. MCDERMOTT, S. ZEGER, and J. SAMET,____ On the use of generalized additive models in time-series studies of air pollution and health, *American Journal of Epidemiology* 156(3): 193-203.

ENGLE R. and C. GRANGER, 1987, Cointegration and error correction: Representation, estimation, and testing, *Econometrica* 55(2): 251-276.

DEMIDENKO, E., 2004, *Mixed Models: Theory and Application*, New Jersey: John Wiley and Sons Inc.

GAO, J., 2007, Nonlinear time series, semiparametric and nonparametric methods, *Monographs on Statistics and Applied Probability* 108, New York: Chapman and Hall.

GAP, J., M. KING, Z. LU, and D. TJØSTHEIM, 2005, Nonparametric specification testing for nonlinear time series with nonstationarity, *Econometric Theory* 25(6): 1869-1892.

HIDALGO, F., 1992, Adaptive semiparametric estimation in the presence of autocorrelation of unknown form, *Journal of Time Series Analysis* 13(1): 47–78.

HASTIE T., and R. TIBSHIRANI, 1990, Generalized additive models, *Monographs on Statistics and Applied probability* 43, London, Chapman and Hall.

LI Q. and J. RACINE, 2006, *Nonparametric Econometrics: Theory and Practice*, New Jersey: Princeton University Press.

MAMMEN, E., O. LINTON and J. NIELSEN, 1999, The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, *Annals of Statistics* 27(5): 1443-1490.

OPSOMER, J., 2000, Asymptotic properties of backfitting estimators, *Journal of Multivariate Analysis* 73(2): 166-179.

SANTOS, E., and E. BARRIOS, 2012, Nonparametric decomposition of time series data with inputs, *Communications in Statistics – Simulation and Computation* 41(9): 1693-1710.

SHIEKE, T., 2001, A generalized additive model for survival times, *Annals of Statistics*, 29(5): 1344-1360.

TJØSTHEIM, D., H. KARLSEN and T. MYKLEBUST, 2007, Nonparametric estimation in a nonlinear cointegration type model, *Annals of Statistics*, 35(1): 252-299.

TRUONG, Y., and C. STONE, 1994, Semiparametric time series regression, *Journal of Time Series Analysis* 15(4): 405–428.

WEI, W., 2006, *Time Series Analysis: Univariate and Multivariate Models* 2nd ed., New York:Addison-Wesley.